International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

PREVENTING PHISHING ATTACK USING EVOLUTIONARY ALGORITHMS

S.DILIP KUMAR[1], A.NAZRATH FATHIMA[2], SIFANA[3], U.JAMRUTH KANI[4]

[1]Assistant Professor, Computer Science and Engineering Department, Arasu Engineering College, Kumbakonam, Tamilnadu, India
[2]UG Scholar, Computer Science and Engineering Department, Arasu Engineering College, Kumbakonam, Tamilnadu, India
[3]UG Scholar, Computer Science and Engineering Department, Arasu Engineering College, Kumbakonam, Tamilnadu, India
[4]UG Scholar Computer Science and Engineering Department, Arasu Engineering College, Kumbakonam, Tamilnadu, India

---

ABSTRACT:

Phishing is a process of fraud activity, in which attacker masquareds as a reputable entity for seizing their user personal details like username, password and other details. It is carried out by Email spoofing or Instant Messaging which insists the user to enter the personal information at a fake websites, the look and feel of the malicious webpages are identical to the legitimate sites. In Existing system, an anti-phishing gateway use uniform resource locator (URL) features and web traffic features to detect phishing websites based on neuro-fuzzy framework. Although features extracted using different dimensions are more comprehensive and these techniques have the drawback as large amount of time for feature extraction and require more training of dataset. To overcome this issue, a search engine is developed using evolutionary algorithms to meet out the performance. The analysis will be carried out by the implication of support vector machine algorithm and proposed system delivers less time consuming, enhance processing speed and it achieves true positive rate in the range of 99% of accuracy.

KEYWORDS: Phishing Website, Machine Learning, SVM, SVM kernel functions.

I. INTRODUCTION:

Phishing is the fraudulent and a criminal activity of acquiring private and sensitive information, such as credit card numbers, personal identification and account usernames and passwords. Using a complex set of social engineering techniques and computer programming expertise, phishing websites gets email recipients and Web users into believing that a spoofed, currently used website is legitimate one. But in actual, the user who is a victim of phishing would realize later that, all his personal and confidential information have been illegally used by stealing it[1]. Due to this activity being performed on a legitimate user account, the privacy and security in using the websites are lost and it offers a wider chance to make use of user details for the criminal purposes by the attacker or the one who steals it. Phishing uses link manipulation, image filter evasion and website forgery to fool Web users. Once the user enters vital information, he immediately becomes a phishing victim and it leads to the illegal use of his personal details.

Thus, for protecting the credentials and preventing it from the fraudsters is inevitable to secure the confidential information. To imply this and to resolve these issues, there is a need to use anti-phishing techniques to secure the details from the attacker who impersonates like the legitimate user. There have been several ways in which phishing is protected. The most common way is to use firewall and anti-virus software. The checking of SSL certificates from websites also gives a hand to identify it to be the legitimate one without any telltale signs to the user. Still, with the emergence of in numerous websites across the world, it is tedious to accurately find and prevent the details from attackers. The anti phishing techniques are well learnt and analyzed for its effective protection against the malicious functions on the networks.[2] We have proposed this paper to overcome all the issues involved in existing algorithms for achieving accurate result and if any phishing is recognised through SVM classification it automatically block the website in which user information are not seized by hackers.

II.RELATED WORK:

2.1:Existing system:

In order to prevent user from accessing phishing sites, there are several approaches have been developed using various algorithm which include Blacklist/white-list method, neuro-fuzzy method, etc. In blacklist/white-list method, set of phishing URLs will be uploaded in blacklist dataset and set of legitimate URLs will be uploaded in white-list dataset. In this technique phishing sites can be recognised and detected by comparing URLs which are uploaded in both dataset. Even though this method has quick detection, managing blacklist/white-list dataset is inefficient due to rapid increase of phishing sites. When the URL of the site is not registered on the URL white-list then the site will be marked as phishing site which leads to achieve high false-positive ratio.

Another approach used in machine learning technique is Fuzzy system. In this method fuzzy rules are established using membership functions. This rule set is not objective and is greatly depend on developer, and the weight value of each main criteria that are used without clarification and the heuristic parameters are very sensitive and difficult to apply in practice due to the complexity of phishing sites.[3] Neural network method is also used for detecting phishing sites but the representation of input to process the algorithm is inefficient and it is lacking in providing information to the training dataset which make false positive ratio to be high[4].

In to improve performance of both the existing Fuzzy and Neural network one has approach a technique named as Neuro-Fuzzy network. This technique is effective to process because all the rule set and membership values are well trained in training dataset and it is easy to integrate the phishing information into neuro-fuzzy network through learning process which enhances the convergence of the training phase[7]. Even though the rules are well trained to detect phishing site, it consumes more time to apply those rules and have less processing speed to detect phishing sites.[4]

2.2:PROPOSED SYSTEM:

In order to overcome the Existing system we have proposed a method using SVM classification. SVM is a supervised Machine learning algorithm which can be mainly used for classification. It contain two phases Training and Testing phase. In trained data set, it efficiently classifies the phishing and legitimate sites by using the kernel functions such as linear , polynomial or sigmoid functions which generate SVM model.[6] In the testing phase, all the features are extracted from test URL. The URL string can be broken down into multiple tokens that constitutes of binary features. Examples of features include length of the URL, number of dots, existence of IP address in the URL and URL with HTTPS and SSL. It also used to check the URL domain name, domain registrar, name server and age of domain. The extracted features are classified based on the training data set. The test URL can be used to predict phishing site based on threshold value calculated from feature vales. Analysis of prediction is done by predefined conditions. When threshold value is greater than zero it make suggestion as phishing sites and block the URL to process it, and if threshold is less than zero it consider as legitimate site and allow the user to access URL.[2] Our proposed system is useful to banking sectors for providing security to user accounts.

III.SYSTEM ARCHITECHTURE:

The proposed architecture is able to detect phishing website effectively by deploying Support Vector Machine techniques.

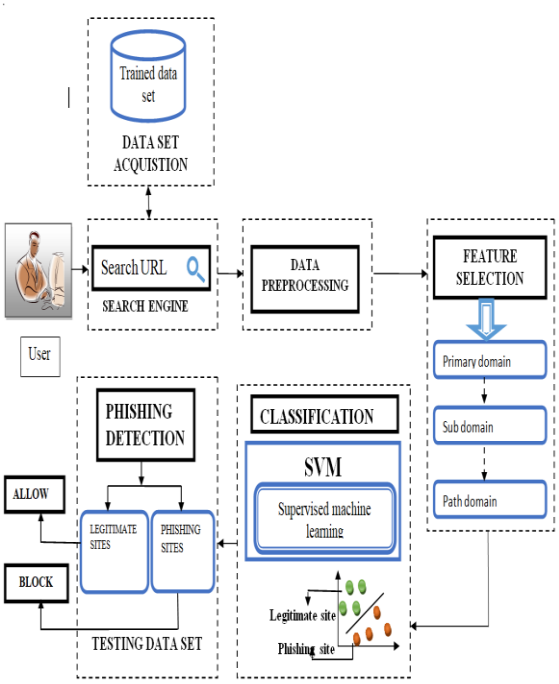


Fig 1. SVM Based Phishing Detection

IV.MODULES DESCRIPTION:

There are five different modules in our system, which are used in our system to detect the phishing sites in our websites. If the site is know to be phishing site, then it will block the user to enter into the sites which protect user from providing their personal information[1].Here we discuss about the detailed description of the modules used in our system.

4.1. DATASET ACQUISTION:

Every data in the world is very important to store which are used to retrieve the better results for the users. So use Data acquisition, it is the process of collecting all the data and uploaded in trained dataset, there will be enough space to store the results to provide accurate results. In data set the classifier is used to classify the phishing site and legitimate sites using some of the kernel functions on the previously collected data[2].

4.2 DATA PREPROCESSING:

After collecting all the data in the dataset, if the user given the test URL then the system will carried out pre-processing. It is a data mining technique. Since Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviours or trends, and is likely to contain many errors. So, It involves transforming of the raw data into an understandable format. The tokenization process will be take place in the test URL such as removing the commas and unwanted attributes in the test URL. It is used for normalizing input URL before feeding it to the algorithm.[3]

4.3. FEATURE SELECTION:

In feature selection there are some feature values which are assigned to the URLs. The collected URLs are transmitted to the feature extractor, which extracts feature values through the predefined URL-based features. The extracted features are stored as input and passed to the classifier generator, which generates a classifier by using input features and the machine learning algorithm. Feature extraction using SVM is to reduce the computational complexity and also increases the classification accuracy of an algorithm.

4.3.1. FEATURES OF URL:

The structure of URL is as follows:
< protocol >: == < SubDomain >: < PrimaryDomain > : < TLD >= < PathDomain >.

The URL:http:// facebook.abc.net/ index.htm includes the following six elements: the protocol is http, the SubDomain is facebook, the PrimaryDomain is abc, the top-level domain (TLD) is net, the Domain is abc.net, and the PathDomain is index.htm. There exist many differences between phishing URLs and legitimate URLs that can be used to recognize easily based on URL features. In particular, we describe in detail three features: the Primary Domain, SubDomain and Path Domain of the URL.[2]

Primary Domain:

Phishers use misspellings or similar Primary Domain of phishing websites to fool users. For example, URL www. facelook.com looks similar to the well-known website www.facebook.com.

Sub Domain:

Phishers often prepend the domain of phishing websites to their website. For example, phishers prepend the SubDomain “facebook.com” to any other domain (e.g., “.io”, “.biz”) that may fool users into the phishing URLs.

Path Domain:

Phishers can use the PathDomain to fool users. For example, phishers may navigate users to the URL www.attack.com/ facebook, where a phishing website interface is similar to the original one. Carelessly, the users will think that this URL is from the “facebook.com” site.

Using all these features, we can identify a phishing website by measuring the similarity score between legitimate and phishing websites. Therefore, we use URL feature for classification to improve the performance.

V.SVM CLASSIFICATION:

This module plays a major role in this system in which the classifier will give the result whether the test URL belongs to a phishing site or the legitimate site by comparing it with the predefined trained dataset. When a page request occurs, the URL of the requested site is transmitted to the feature extractor, which extracts the feature values through the predefined URL-based features. Then the comparision will be held by comparing its threshold value assigned by the feature extractor.

PHISHING WEBSITE DETECTION:

The phishing site is identified based on learned information from training dataset and then alerts the page-requesting user about the SVM classification result. If the page is a phishing it will block the user to enter into the site, otherwise it will allow the user to access the site.[4] In such a way our proposed system prevents the user from losing their personal information.

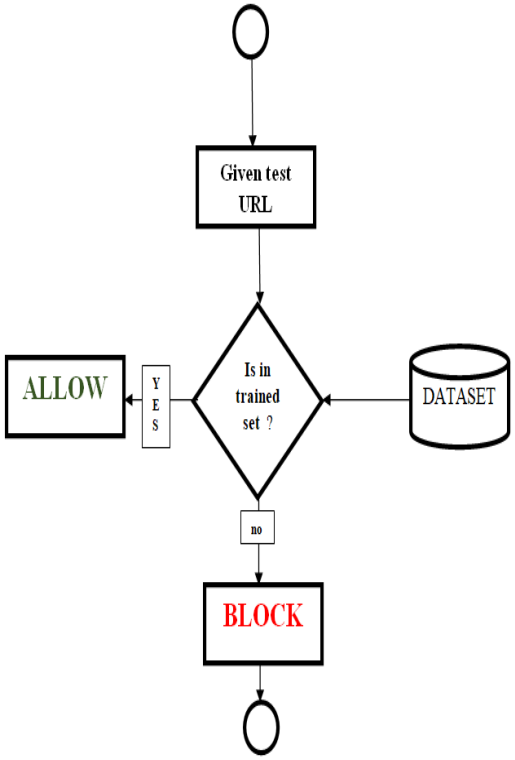


Fig 2. Detection Phase

ALGORITHM 1:

- Input: X is a DomainDataset
- Output: The heuristic value of Domain

```
if X=0 then
Return X belongs to a legitimate site ;
else Result = SuggestionGoogle(X);
if Result is NULL then
 Return X belongs to a legitimate site;
else value = Search(Result,X);
Return value;
end
end
```

ALGORITHM 2:

Input : D is a Trained Data Set , X is a Test URL.

Output : Return the result of normal or phishing sites.

If X in D set **then**

//legitimate site detection

Return x is a legitimate site,

ALLOW to enter into the site

Else

//phishing site detection

Return X is a phishing site,

BLOCK the user to enter into the site

End

SYSTEM REQUIREMENTS:

HARDWARE REQUIREMENTS:

Processor : Quad core processor

RAM : 4 GB

Hard disk : 500 GB

Input device : Standard keyboard and 15 inch color monitor.

SOFTWARE REQUIREMENTS:

Operating System : Windows 10

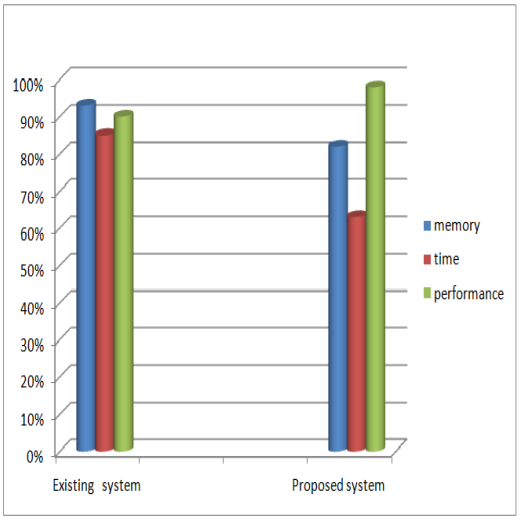
Front End : PHP

IDE :Macromedia Dreamweaver 8

Back End : MYSQL SERVER

PERFORMANCE ANALYSIS:

In this section, the performance evaluation of this work is done to prove the performance improvement over the proposed methodology than the existing system in terms of time, memory usage and quality of generated rules with maximizing the fitness function. The proposed application concentrates on the SVM classification based on the threshold value derived from feature values.



	Memory	Time	Performance
Existing system	93%	85%	90%
Proposed system	82%	63%	98%

Chart 1: Efficiency of the proposed Method

CONCLUSION:

Detecting the malicious URL is one of the crucial problems in the internet. In this paper, we propose an SVM-based phishing URL detection solution. Support Vector Machine algorithm achieved high classification accuracy over existing neuro-fuzzy network. Our proposed method has been implemented on dataset of phishing and legitimate

URLs. The set URLs are analyzed using inter related features and the experiment furnishes a classification of phishing and legitimate URL with 97.83% of accuracy and 1.82% of false positive rate. Thus, this approach is based on an supervised machine learning technique that rely on characteristics of phishing URL properties to detect and prevent phishing website and to provide high level security based on SVM classification.

REFERENCE:

[1] Automated Phishing Website Detection Using URL Features and Machine Learning Technique V.Preethi, G.Velmayil *oaji.net/pdf.html?n=2017/1992-1514529507* ln October 2016

[2] Phishing-Aware: A Neuro-Fuzzy Approach for Anti-Phishing on Fog Networks Chuan Pham_x, Luong A. T. Nguyenz, Nguyen H. Tran, Eui-Nam Huh, Choong Seon Hong 2018 https://ieeexplore.ieee.org/document/8352739/

[3]Prakash P., Manish K., Kompella R.R., Gupta M., "PhishNet: Predictive Blacklisting to Detect Phishing Attacks", presented as part of the Mini-Conference at IEEE INFOCOM 2016

[4] Extraction of Features and Classification on Phishing Websites using Web Mining Techniques Nandhini.S , Dr.V.Vasanthi M.Phil. Scholar. https://www.ijedr.org/papers/IJEDR1704198.pdf on 2017

[5] Phishing Website Detection based on Multidimensional Features driven by Deep Learning Peng Yang, Guangzhen Zhao, Peng Zeng. https://www.researchgate.net/.../279921715

[6]. Phishing Website Detection based on Multidimensional Features driven by Deep Learning Peng Yang, Guangzhen Zhao, Peng Zeng. https://www.researchgate.net/.../330326876 on jan2019.

[7]. Unethical Network Attack Detection and Prevention using Fuzzy based Decision System in Mobile Ad-hoc Networks. R.Thanuja,J Electr Eng Technol.2018; 13(5): 2086-2098. http://doi.org/10.5370/JEET.2018.13.5.2086